Enhanced convergence estimates for semi-lagrangian schemes Application to the Vlasov-Poisson equation

Frédérique Charles, Bruno Després, Michel Mehrenberger

Enhanced convergence estimates for semi-lagrangian schemes Application to the Vlasov-Poisson equation

Frédérique Charles, Bruno Després, Michel Mehrenberger

October 5, 2011

Abstract

We prove enhanced error estimates for high order semi-lagrangian discretizations of the Vlasov-Poisson equation. It provides new insights into optimal numerical strategies for the numerical solution of this problem. The new error estimate

$$O\left(\min\left(\frac{\Delta x}{\Delta t},1\right)\Delta x^p+\Delta t^2\right)$$

is based on advanced error estimates for semi-lagrangian schemes, also equal to shifted Strang's schemes, for the discretization of the advection equation.

1 Introduction

The aim in this work is to prove enhanced error estimates for high order semi-lagrangian discretizations of the Vlasov-Poisson equation [12]. In this work enhanced means two things. First it means enhanced with respect to the time step Δt, that is we do not want the error estimate to be spoiled by a $\frac{1}{\Delta t}$ which is very often encountered in the numerical analysis of the Vlasov-Poisson equation as in [6, 1, 10]. Second, in the case of advection equation in dimension one, it means enhanced with respect to the norm, that is the error and the regularity of the solution are evaluated in the same norm.

Our main result is stated in Theorem 1. Estimate (6) has the advantage that it is non singular for small Δt. It is a strong improvement with respect to the literature [6, 1, 10]. This new error estimate is based on new estimates for the advection equation on a regular periodic grid for which we use sharp properties of semi-lagrangian schemes also equal to shifted Strang's stencil [14, 8, 9], see also [2, 3, 4, 5]. We will make use of the connection of these numerical schemes with B-Splines techniques [11, 13].

Theorem 1. *Consider the Vlasov-Poisson equation*

$$\frac{\partial f}{\partial v} + v\frac{\partial f}{\partial x} + E(t,x)\frac{\partial f}{\partial v} = 0, \tag{1}$$

with

$$\frac{\partial E}{\partial x}(t,x) = \int_{-\infty}^{+\infty} f(t,x,v)dv - 1, \tag{2}$$

where $f(t,x,v)$ is the distribution function of charged particles (ions or electrons) and $E(t,x)$ the self-consistent electric field. We consider periodic boundary conditions on the variable $x \in [0,1]$, that is

$$f(t,0,v) = f(t,L,v), \quad v \in \mathbb{R}, \quad t \geq 0, \tag{3}$$

and

$$E(t,0) = E(t,L). \tag{4}$$

Assume the exact solution is sufficiently smooth and has compact support. Consider a semi-lagrangian scheme of order p direction by direction for the discretization of (63-66) with Strang's splitting in time. Assume the grid is such that $\Delta v = \alpha \Delta x$ is proportional to Δx, α kept constant. Assume $n\Delta t \leq T$.

There exists a constant $C > 0$ which depends on T, on the regularity of the solution and of the parameters of the problem such that the numerical error $e^n = f^n - f(t_n)$ is bounded in the L^2 norm[1] *by*

$$||e^n||_{L^2} \leq C\left(\min\left(\frac{\Delta x}{\Delta t}, 1\right)\Delta x^p + \Delta t^2\right). \tag{6}$$

The organization of this work is as follows. First in Section 2 we detail optimal numerical strategies which are immediate consequence of the error estimate (6). The next section is devoted to the design of enhanced error estimates for

[1] We will use natural notations for the norm of discrete functions over the real line $\mathbb{R}$ discretized with a mesh length $\Delta x > 0$. For example the L^p norm of a discrete function $w = (w_i)_{i\in\mathbb{Z}}$ defined over the entire real line is

$$\|w\|_p = \left(\Delta x \sum_i |w_i|^p\right)^{\frac{1}{p}} \quad 1 \leq p < \infty, \quad \text{and } \|w\|_\infty = \sup_i |w_i|.$$

If the domain is finite, for example $\Omega =]0,1[^2$, then $N\Delta x = 1$ is required for some $N \in \mathbb{N}$: in this case the discrete function is $w = (w_{ij})_{1\leq i,j\leq N}$ with norms defined by

$$\|w\|_p = \left(\Delta x \sum_{1\leq i,j\leq N} |w_{ij}|^p\right)^{\frac{1}{p}} \quad 1 \leq p < \infty, \quad \text{and } \|w\|_\infty = \sup_{1\leq i,j\leq N} |w_{ij}|. \tag{5}$$

These notations are compatible with the standard definition of the L^p norm of a function

$$\|z\|_{L^p(\Omega)} = \left(\int_\Omega |z(x)|^p dx\right)^{\frac{1}{p}} \quad 1 \leq p < \infty, \quad \text{and } \|z\|_{L^\infty(\Omega)} \sup_{x\in\Omega} |z(x)|.$$

the numerical discretization of the advection equation in dimension one with semi-lagrangian schemes of arbitrary orders (also equal to shifted Strang's stencils). After that we prove the main theorem. In the appendix we provide the reader with more advanced formulas for the numerical approximation of the advection equation.

2 Application

The computation of a numerical solution to the Vlasov-Poisson equation is very heavy task from the CPU point of view. It is therefore of major interest to study optimal numerical strategies such the upper bound (6) of the error is as small as possible as a function of Δx and Δt. We consider three different strategies with $p \geq 3$ since it corresponds to the situation we are interested in. It will appear that the fact that singularity for very small Δt is removed in the right hand side of (6) due to the term $\min\left(\frac{\Delta x}{\Delta t}, 1\right)$ has an immediate consequence on the optimal scaling laws. Essentially we will obtain that

$$\Delta t \approx \Delta x^{\frac{p}{2}} << \Delta x$$

which yields that $\min\left(\frac{\Delta x}{\Delta t}, 1\right) = 1$. Note that the results would not be the same, if we would consider higher order discretizations in time (see [6], for a discussion).

2.0.1 Minimizing the error at given storage

The storage is the memory requirement needed to run a given computation. Since the storage is proportional to $1/\Delta x \Delta v \approx 1/\Delta x^2$, it means that we look for an optimal time step Δt which minimizes the error at a given (but small) Δx.
So we are looking for a scaling law

$$\Delta t = \Delta x^{\beta}, \qquad \beta > 0$$

such that the right hand side (6) is as small as possible when Δx goes to zero. Since $\min\left(\frac{\Delta x}{\Delta t}, 1\right) \Delta x^p + \Delta t^2 \leq \Delta x^p + \Delta t^2$ it is immediate that $\beta \geq \frac{p}{2} \geq \frac{3}{2} > 1$. In this case $\min\left(\frac{\Delta x}{\Delta t}, 1\right) = \min\left(\frac{\Delta x}{\Delta x^{\beta}}, 1\right) = 1$ for small Δx. We obtain asymptotically $||e^n||_{L^2} \leq C\Delta x^p$ with a larger constant. Our interest being to have nevertheless the biggest time step to minimize the overall cost of the computation we obtain the scaling law

$$\beta = \frac{p}{2}.$$

2.0.2 Minimizing the error at given CPU cost

The CPU cost is proportional to the total number of cells and to the number of time steps, that is after normalization

$$\text{CPU} \approx \frac{1}{\Delta t \Delta x^2}.$$

Therefore $\Delta t = \frac{1}{\alpha \Delta x^2}$ where $\alpha >> 1$ is the normalized numerical value of the CPU cost. Plugging in the error formula (6) we find

$$\text{Error} \approx C \left(\min(\alpha \Delta x^3, 1) \Delta x^p + \frac{1}{\alpha^2 \Delta x^4} \right).$$

The minimum is approximatively obtained by equating the two contributions

$$1 = \min(\alpha \Delta x^3, 1) \left(\alpha \Delta x^3 \right)^2 \Delta x^{p-2}.$$

Therefore $1 \leq \alpha^2 \Delta x^{p+4}$ and $\Delta x \geq \alpha^{-\frac{2}{p+4}}$. So $\alpha \Delta x^3 \geq \alpha^{1-\frac{6}{p+4}} \geq \alpha^{1-\frac{6}{3+4}} >> 1$ since $p \geq 3$ by hypothesis. So $\min(\alpha \Delta x^3, 1) = 1$. In conclusion the optimal strategy that minimizes the error at given CPU cost proportional to α is

$$\Delta x \approx \alpha^{-\frac{2}{p+4}} \text{ and } \Delta t \approx \alpha^{-1+\frac{4}{p+4}} = \alpha^{-\frac{p}{p+4}}.$$

We observe $\Delta t \approx \Delta x^{\frac{p}{2}}$ that for large p: this is the scaling law of the previous strategy.

2.0.3 Minimizing the CPU cost a given error

The CPU cost is a strictly convex function with respect to $(\Delta t, \Delta x)$. Assume the error ε is very small, that is

$$\min \left(\frac{\Delta x}{\Delta t}, 1 \right) \Delta x^p + \Delta t^2 = \varepsilon, \quad \text{with } \varepsilon << 1.$$

We formulate the problem as a minimum problem with a constraint. We distinguish two cases.

First case: $\Delta t > \Delta x$: Minimum solutions, if they exist, are the critical point of the Lagrangian

$$L = \frac{1}{\Delta t \Delta x^2} - \lambda \left(\frac{\Delta x^{p+1}}{\Delta t} + \Delta t^2 - \varepsilon \right)$$

where λ is the Lagrange multiplier. The optimality conditions are

$$\begin{cases} \partial_{\Delta t} L = -\frac{1}{\Delta t^2 \Delta x^2} - \lambda \left(-\frac{\Delta x^{p+1}}{\Delta t^2} + 2\Delta t \right) = 0, \\ \partial_{\Delta x} L = -\frac{2}{\Delta t \Delta x^3} - \lambda (p+1) \frac{\Delta x^p}{\Delta t} = 0. \end{cases}$$

Since $p \geq 3$ one may approximate

$$-\frac{\Delta x^{p+1}}{\Delta t^2} + 2\Delta t \approx 2\Delta t$$

for small Δx. Therefore the optimality conditions imply

$$\begin{cases} -\frac{1}{\Delta t^2 \Delta x^2} \approx \lambda 2 \Delta t, \\ -\frac{2}{\Delta t \Delta x^3} = \lambda (p+1) \frac{\Delta x^p}{\Delta t}. \end{cases}$$

It yields

$$\frac{2\Delta t}{\Delta x} \approx \frac{p+1}{2}\frac{\Delta x^p}{\Delta t^2}$$

that is

$$\Delta t \approx \left(\frac{p+1}{4}\right)^{\frac{1}{3}} \Delta x^{\frac{p+1}{3}} << \Delta x$$

since $p \geq 3$. It is in contradiction with the hypothesis $\Delta t < \Delta x$. It means that there is no solution in this case.

Second case: $\Delta t \leq \Delta x$: The Lagrangian is

$$L = \frac{1}{\Delta t \Delta x^2} - \lambda \left(\Delta x^p + \Delta t^2 - \varepsilon\right).$$

The optimality conditions are

$$\begin{cases} \partial_{\Delta t} L = -\frac{1}{\Delta t^2 \Delta x^2} - \lambda 2\Delta t = 0, \\ \partial_{\Delta x} L = -\frac{2}{\Delta t \Delta x^3} - \lambda p \Delta x^{p-1} = 0 \end{cases}$$

whose solution is given by

$$\Delta t = \left(\frac{p+1}{4}\right)^{\frac{1}{2}} \Delta x^{\frac{p}{2}} << \Delta x.$$

Once again this scaling law is very close to the first one.

2.0.4 Exponential integrators in time

Suppose that for each splitting step, instead of solving over Δt the $1D$ advection equation, one solves N times the same equation with time step $\Delta t/N$, and that we consider the limit as N goes to infinity of such scheme. In the previous analysis [10], we cannot conclude that the scheme converges. On the contrary, in our new framework, one obtains the rate

$$O(\Delta x^p + \Delta t^2).$$

3 A pedagogical example for the advection equation

We consider in this section the numerical discretization of the advection equation with initial condition

$$\begin{cases} \partial_t u + v\partial_x u = 0, \qquad v > 0, \\ u(0,x) = u_0(x) \end{cases} \tag{7}$$

in one dimension by means of a semi-lagrangian scheme. We recall that the solution of problem (7) verifies

$$u(t+s,x) = u(t,x-vs) \quad \forall t \geq 0,\, s \geq 0. \tag{8}$$

We introduce a Cartesian regular discretization $\{x_i\}_{i\in\mathbb{Z}} = \{i\Delta x\}_{i\in\mathbb{Z}}$ of $\mathbb{R}$ and a time step Δt. For the sake of simplicity the numerical initial condition is always exact, that is

$$u_j^0 = u_0(x_j) \qquad \forall j \in \mathbb{Z}. \tag{9}$$

Assuming that we know the approximation of the solution of (7) at the time $t^n = n\Delta t$ on the mesh, we set for all i

$$u_i^{n+1} = \tilde{u}_i^n\left(x_i - v\Delta t\right) \tag{10}$$

where $\tilde{u}^n$ is a continuous function obtain by means of an interpolation of values $\{u_i^n\}_{i\in\{k_1,\ldots,k_2\}}$. Such a procedure is called semi-lagrangian because the discrete solution is given on a eulerian fixed mesh, but on the other hand the scheme relies on the construction of $\tilde{u}^n\left(x_i - v\Delta t\right)$ which is used in the lagrangian approximation (10). This is described in Figure 1.

In order to fully understand the main difficulty about the design of optimal estimates, we consider, in a first part for pedagogical purposes, a second order lagrangian interpolation (that is $k_2 = k_1 + 1$). It will enlighten the interest of having an interpretation of such scheme both in terms of a semi-lagrangian interpolation and in terms of a standard numerical approximation of a partial differential equation. Then we will develop both approaches for general high order semi-lagrangian schemes.

3.0.5 Numerical scheme

The function $\tilde{u}^n$ is obtained by a lagrangian interpolation from the two neighboring points. Let $x_i - v\Delta t$ be the foot of the characteristics and $r \in \mathbb{Z}$ such that

$$x_{i+r} \leq x_i - v\Delta t < x_{i+r+1},$$

that is $r \leq -v\frac{\Delta t}{\Delta x} < r+1$. In other terms we set

$$r = E\left(-v\frac{\Delta t}{\Delta x}\right) \leq -1. \tag{11}$$

Since $v > 0$ then $r + 1 \leq 0$ for any $\Delta t > 0$ and $\Delta x > 0$. This is illustrated in Figure 1.

The second order reconstruction of $\tilde{u}^n$ on $[x_{i+r}, x_{i+r+1}[$ is

$$\tilde{u}_i^n(x) = \frac{x_{i+r+1} - (x_i - v\Delta t)}{\Delta x} u_{i+r}^n + \frac{(x_i - v\Delta t) - x_{i+r}}{\Delta x} u_{i+r+1}^n. \tag{12}$$

Introducing this expression in (10) one obtains

$$u_i^{n+1} = \nu u_{i+r}^n + (1-\nu) u_{i+r+1}^n, \tag{13}$$

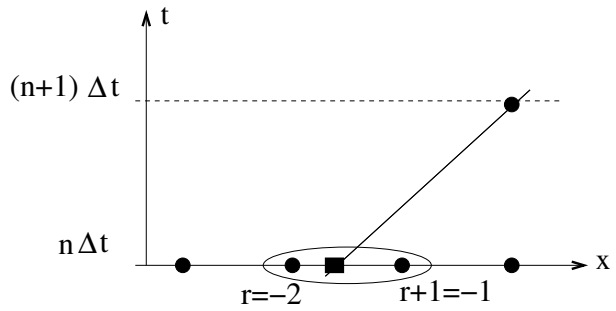

Figure 1: Stencil of the second order lagrangian ($p+1=2$) interpolant with a shift of one cell: here $r=-2$.

where we set

$$\nu := r + 1 + \frac{v\Delta t}{\Delta x}. \tag{14}$$

We deduce from the definitions of r (11) that

$$0 < \nu \leq 1. \tag{15}$$

The standard way to evaluate the error in space attached to this procedure is the following. First the interpolation error of (12) is $O(\Delta x^2)$ for a smooth function. Second it is without to say that (13) is stable in the maximum norm. So one just sums the errors from one time step to the following and obtain

$$\max_i |u(n\Delta t, x_i) - u_i^n| \leq n\, O(\Delta x^2) \leq CT\frac{\Delta x^2}{\Delta t}, \qquad n\Delta t \leq T. \tag{16}$$

The constant C is independent of Δt, Δx and v. The estimate (16) is second order in space but is clearly non optimal for small Δt. Indeed (16) corresponds to the upwind scheme for $r-1$: in this case it is well known that the error can be bounded by $O(\Delta x)$. This analysis enlightens the fact that a naive interpolation error estimate is not optimal in our context.

3.0.6 The truncation error

We now desire to explain how to recast the previous analysis of the interpolation error so as to obtain optimal estimate for small Δt. An idea is to rewrite (16) under the form of the Finite Difference scheme but with skewed discrete derivatives

$$\frac{u_i^{n+1} - u_{i+r+1}^n}{\Delta t} + \left(v + \frac{(r+1)\Delta x}{\Delta t}\right)\frac{u_{i+r+1}^n - u_{i+r}^n}{\Delta x} = 0. \tag{17}$$

It is usual to introduce the truncation error $\varepsilon^n = (\varepsilon_i^n)_{i\in\mathbb{Z}}$

$$\begin{aligned}\varepsilon_i^n =& \frac{u\left((n+1)\Delta t, x_i\right) - u\left(n\Delta t, x_{i+r+1}\right)}{\Delta t} \\ &+ \left(v + \frac{(r+1)\Delta x}{\Delta t}\right) \frac{u\left(n\Delta t, x_{i+r+1}\right) - u\left(n\Delta t, x_{i+r}\right)}{\Delta x},\end{aligned} \tag{18}$$

where u is the solution of equation (7). The next task consists is showing that it can be related to a more standard truncation error.

The unit time $\frac{\Delta x}{v}$ is the one needed for a characteristics to travel in a cell of length Δx, from one side to the other. Let $\Delta s \geq 0$ be defined as Δt minus a $d\frac{\Delta x}{v}$ where $d+1$ is the largest as possible integer. A convenient definition of Δs is

$$\Delta s := (r+1)\frac{\Delta x}{v} + \Delta t. \tag{19}$$

We deduce from the definition of r (11) and ν (14) :

$$0 < \Delta s \leq \Delta t, \quad \nu = \frac{v\Delta s}{\Delta x}. \tag{20}$$

Let $\bar{\varepsilon}$ be the truncation error of the same scheme but with the time step Δs

$$\begin{aligned}\bar{\varepsilon}_{i+r+1}^n =& \frac{u\left(n\Delta t + \Delta s, x_{i+r+1}\right) - u\left(n\Delta t, x_{i+r+1}\right)}{\Delta s} \\ &+ v\frac{u\left(n\Delta t, x_{i+r+1}\right) - u\left(n\Delta t, x_{i+r}\right)}{\Delta x}.\end{aligned} \tag{21}$$

Proposition 2. *One has the formula* $\varepsilon_i^n = \frac{\Delta s}{\Delta t}\,\bar{\varepsilon}_{i+r+1}^n$.

Proof. The exact solution is constant along the characteristics, that is

$$u\left((n+1)\Delta t, x_i\right) = u\left(n\Delta t + \Delta s, x_{i+r+1}\right).$$

We use this relation in (18) to eliminate $u\left((n+1)\Delta t, x_i\right)$. The claim is evident on the resulting quantity. □

Proposition 3. *Assume that* $u_0 \in W^{2,\infty}(\mathbb{R})$. *There is a constant* $C > 0$, *independent of* Δt, Δx *and* v, *such that*

$$\left\|\bar{\varepsilon}^n\right\|_\infty \leq Cv\left\|u_0''\right\|_{L^\infty(\mathbb{R})}\Delta x. \tag{22}$$

Proof. A standard Taylor-Lagrange expansion on the error $\bar{\varepsilon}_i^n$ gives

$$\begin{aligned}\bar{\varepsilon}_i^n =& \left(\partial_t u + v\partial_x u\right)(n\Delta t, x_i) \\ &+ \frac{\Delta s}{2}\partial_{tt}u(n\Delta t + \tau\Delta s, x_i) + \frac{v\Delta x}{2}\partial_{xx}u(n\Delta t, x_i + \xi\Delta x)\end{aligned}$$

and then

$$\begin{aligned}\left\|\overline{\varepsilon}^n\right\|_\infty &= O\left(\Delta s \left\|\partial_{tt} u\right\|_{L^\infty(\mathbb{R})}\right) + O\left(v\ \Delta x \left\|\partial_{xx} u\right\|_{L^\infty(\mathbb{R})}\right) \\ &= O\left(v^2 \Delta s \left\|\partial_{xx} u\right\|_{L^\infty(\mathbb{R})}\right) + O\left(v\ \Delta x \left\|\partial_{xx} u\right\|_{L^\infty(\mathbb{R})}\right) \\ &= O\left((1+\nu) v \Delta x \left\|u_0''\right\|_{L^\infty(\mathbb{R})}\right)\end{aligned}$$

thanks to (20). Using (15) it proves the claim. □

3.0.7 An optimal error estimate

The numerical error is $e^n = (e_i^n)_{i\in\mathbb{Z}}$. with $e_i^n = u(n\Delta t, x_i) - u_i^n$.

Proposition 4. *One has the error estimate for all n such that $n\Delta t \leq T$*

$$\left\|e^n\right\|_\infty \leq \left(CT \left\|u_0''\right\|_{L^\infty(\mathbb{R})}\right) \frac{v\Delta s \Delta x}{\Delta t}. \tag{23}$$

The constant $C > 0$ is independent of Δt, Δx and v.

Proof. Let us denote the iteration operator $\mathcal{R}_\nu$ so that (13)-(17) is rewritten as $u^{n+1} = \mathcal{R}_\nu u^n$. Thanks to (21) one has

$$u\left((n+1)\Delta t, \cdot\right) = \mathcal{R}_\nu u\left(n\Delta t, \cdot\right) + \Delta t\, \varepsilon^n.$$

Therefore the error e^n satisfies $e^{n+1} = \mathcal{R}_\nu e^n + \Delta t\, \varepsilon^n$. Since the numerical scheme satisfies the maximum principle, then $\left\|\mathcal{R}_\nu\right\|_\infty \leq 1$. Therefore

$$\left\|e^{n+1}\right\|_\infty \leq \left\|e^n\right\|_\infty + \Delta t \left\|\varepsilon^n\right\|_\infty.$$

Summation over n yields $\left\|e^n\right\|_\infty \leq n\Delta t \frac{v\Delta s}{\Delta t} \left\|\overline{\varepsilon}^n\right\|_\infty$. Using (22) it proves the claim. □

Remark 5. *This error estimate is now optimal with respect to Δt. If $\Delta s = \Delta t$ which means that the characteristics does not go out the first cell, then we recover the first order estimate of convergence characteristics of the upwind scheme. If $\Delta s << \Delta t$ one can use the bound $v\Delta s \leq \Delta x$ which is always true and recover the standard error estimate obtained in the numerical analysis of semi-lagrangian numerical methods. We also notice that the error estimate is optimal with respect to v since the error vanishes if $v \to 0^+$.*

Remark 6. *The estimate (23) can be enhanced with a term $1-\nu$ in the right hand side. This is due to the fact that the scheme is exact if $\nu = 1$. This term is visible in the general estimate (40).*

4 Semi-lagrangian interpolation schemes of order $p+1$

This section is devoted to the study of arbitrary order semi-lagrangian schemes.

4.1 Numerical scheme

In this section the interpolation function $\tilde{u}^n$ is obtained by a lagrangian interpolation of $p+1$ points. Like in previous section, we introduce $r \in \mathbb{Z}$ such that $x_{i+r} \leq x_i - v\Delta t < x_{i+r+1}$. and we introduce also $k \in \mathbb{Z}$: the expression of $\tilde{u}^n$ on $[x_{i+r+1+k-p}, x_{i+r+1+k}]$ is obtained with a lagrangian interpolation of the values $\{u_{i+r+1+k-p}, \ldots, u_{i+r+1+k}\}$

$$\tilde{u}^n(x) = \sum_{l=k-p}^{k} L_{i+r+l}(x) u^n_{i+r+1+l} \tag{24}$$

where the Lagrange polynomials are $L_{i+r+1+l}(x) = \frac{\prod_{h=k-p,h\neq l}^{k}(x - x_{i+r+1+h})}{\prod_{h=k-p,h\neq l}^{k}(x_{i+r+l} - x_{i+r+1+h})}$.

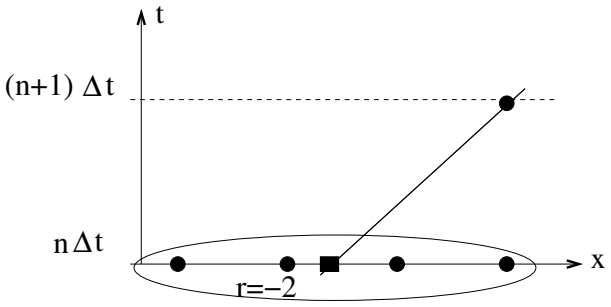

Figure 2: Example of a semi-lagrangian scheme: here $r = -2$, $k = 1$ and $p = 3$.

We define for convenience

$$\alpha_l(\nu, k, p) = L_{i+r+1+l}\left(x_i - v\Delta t\right) = \frac{\prod_{h=k-p,h\neq l}^{k}\left(h + \nu - 1\right)}{\prod_{h=k-p,h\neq l}^{k}\left(h - l\right)}.$$

The reduced Courant number is $\nu = \frac{v\Delta s}{\Delta x} \in]0, 1]$ and $\Delta s = (r+1)\frac{\Delta x}{v} + \Delta t$ as before. The scheme can be written as

$$u_i^{n+1} = \sum_{l=k-p}^{k} \alpha_l\left(\nu, k, p\right) u^n_{i+r+1+l}, \qquad i \in \mathbb{Z}. \tag{25}$$

Four parameters characterize this formula: $p+1$ is the interpolation order, k determines the local stencil shift and ν is the reduced Courant number; the number of cells that are crossed by the characteristics is (11) is r. The iteration operator is conveniently defined as $\mathcal{R}_{\nu,k,p}$ so that (25) is equivalent to

$$u^{n+1} = \mathcal{R}_{\nu,k,p} u^n. \tag{26}$$

4.2 Stability properties

Introducing the discrete Fourier transform of the vector u^n

$$\bar{u}^n(\psi) = \sum_{j=-\infty}^{\infty} e^{-\mathbf{i}jy\psi} u_j^n, \qquad \mathbf{i}^2 = -1, \tag{27}$$

we get from (25)

$$\bar{u}^{n+1}(\psi) = \lambda_{\nu,k,p}(\psi)\bar{u}^n(\psi), \tag{28}$$

where

$$\lambda_{\nu,k,p}(\psi) = \sum_{l=k-p}^{k} \alpha_l(\nu,k,p)\, e^{\mathbf{i}\,(r+1+l)\psi} \tag{29}$$

is the amplification factor for Fourier modes of this scheme. We also introduce the reduced amplification factor is given by

$$\mu_{\nu,k,p}(\psi) = \sum_{l=k-p}^{k} \alpha_l(\nu,k,p)\, e^{\mathbf{i}\,l\psi}, \tag{30}$$

and we have the relation

$$\lambda_{\nu,k,p}(\psi) = e^{\mathbf{i}(r+1)\psi}\mu_{\nu,p,k}(\psi). \tag{31}$$

Proposition 7. *The amplification factor satisfies*

$$|\lambda_{\nu,k,p}(\psi)| \leq 1, \qquad \forall \psi \in \mathbb{R} \tag{32}$$

if and only if $p \in \{2k, 2k+1, 2k+2\}$.

Many proves are available in the case $r+1=0$. We refer to the seminal work of Strang in another context [14, 8, 9]. See also [5, 10, 7]. In consequence one has the following stability property.

Proposition 8. *Assume that* $p \in \{2k, 2k+1, 2k+2\}$. *The iteration operator is bounded as an operator over discrete functions with bounded* L^2 *norm. More precisely*

$$\|\mathcal{R}_{\nu,k,p}u\|_2 \leq \|u\|_2, \quad \forall u = (u_j)_{j\in\mathbb{Z}}. \tag{33}$$

4.3 Interpolation-based error estimate in L^2

We introduce the numerical error of the scheme defined by $e_i^n := u(n\Delta t, x_i) - u_i^n$ and $e^n = (e_i^n)_{i\in\mathbb{Z}}$. One has the standard decomposition

$$e^{n+1} = \mathcal{R}_{\nu,k,p}\; e^n + g^n \tag{34}$$

where $g^n = (g_i^n)$ is the interpolation error

$$g_i^n = u\left(n\Delta t, x_i - v\Delta t\right) - \sum_{l=k-p}^{k} \alpha_l(\nu,k,p) u\left(n\Delta t, x_{i+r+1+l}\right).$$

We will use the following standard interpolation formula for which we refer the reader to [13] page 124. Let Q_i^{p+1} be the spline function over the $p+2$ points $x_i - v\Delta t$ and $x_{i+r+1+l}$ for $l = k-p \leq i \leq k$. Q_i^{p+1} is a piecewise polynomial of degree n and has compact support in $]x_{i+r+1+k-p}, x_{i+r+1+k}[$ since $x_i - v\Delta t \in]x_{i+r+1+k-p}, x_{i+r+1+k}[$ by definition of r. The interpolation formula writes

$$g_i^n = \frac{\omega_i}{p!} \int_{x_{i+r+1+k-p}}^{x_{i+r+1+k}} Q_i^{p+1}(t) u^{(p+1)}(t) dt \tag{35}$$

with

$$\omega_i = \prod_{l=i+r+1+k-p}^{l=i+r+1+k} (x_i - v\Delta t - x_l).$$

Other properties of Q_i^{p+1} are ([13] page 124)

$$\int_{x_{i+r+1+k-p}}^{x_{i+r+1+k}} Q_i^{p+1}(t) dt = \frac{1}{p+1}, \tag{36}$$

and

$$0 \leq Q_i^{p+1}(t) \leq \frac{1}{x_{i+r+1+k} - x_{i+r+1+k-p}} = \frac{1}{p\Delta x}. \tag{37}$$

Proposition 9. *One has the inequality*

$$\|g^n\|_2 \leq C_{k,p} \nu(1-\nu) \left\| u_0^{(p+1)} \right\|_{L^2(\mathbb{R})} \Delta x^{p+1} \tag{38}$$

where

$$C_{k,p} = \frac{1}{(p+1)^{\frac{1}{2}}} \times \begin{cases} \frac{(k+1)!k!}{(2k+1)!} & p = 2k+1, \\ \frac{(k)!(k)!}{(2k)!} & p = 2k, \\ \frac{(k+1)!(k+1)!}{(2k+2)!} & p = 2k+2. \end{cases}$$

Proof. First we use (19) to eliminate $v\Delta t = v\Delta s - (r+1)\Delta x$ in the interpolation formula (35). We obtain

$$\omega_i = \prod_{l=k-p}^{k} (-v\Delta s - l\Delta x) = (-1)^{p+1} \Delta x^{p+1} \prod_{l=k-p}^{k} (l+\nu),$$

so

$$|\omega_i| \leq \Delta x^{p+1} \left| \prod_{l=k-p}^{k} (l+\nu) \right|.$$

Consider for example the case $p = 2k+1$. A rearrangement shows that

$$\begin{aligned} \left| \prod_{l=k-p}^{k} (l+\nu) \right| &= \nu(1-\nu) \underbrace{(1+\nu)(2-\nu)}_{\leq 1\times 2} \cdots \underbrace{(k+\nu)(k+1-\nu)}_{\leq k \times k+1} \\ &\leq \nu(1-\nu) \left(1 \times \cdots \times k\right) \left(2 \times \cdots \times (k+1)\right) \\ &\leq \nu(1-\nu) p! (p+1)^{\frac{1}{2}} C_{k,p}. \end{aligned} \tag{39}$$

A similar trick shows that $\left|\Pi_{l=k-p}^{k}(l+\nu)\right| \leq \nu(1-\nu)p!C_{k,p}(p+1)^{\frac{1}{2}}$ in all cases.

Secondly we set $I_i =]x_{i+r+1+k-p}, x_{i+r+1+k}[$. The Cauchy-Schwarz inequality applied to (35) yields

$$\begin{aligned}
|g_i^n| &\leq \frac{|\omega_i|}{p!}\left\|Q_i^{p+1}\right\|_{L^2}\left\|u^{(p+1)}\right\|_{L^2(I_i)} \\
&\leq \frac{|\omega_i|}{p!}\left\|Q_i^{p+1}\right\|_{L^1}^{\frac{1}{2}}\left\|Q_i^{p+1}\right\|_{L^\infty}^{\frac{1}{2}}\left\|u^{(p+1)}\right\|_{L^2(I_i)} \\
&\leq \frac{|\omega_i|}{p!}\left(\frac{1}{p+1}\right)^{\frac{1}{2}}\left(\frac{1}{p\Delta x}\right)^{\frac{1}{2}}\left\|u^{(p+1)}\right\|_{L^2(I_i)}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\|g^n\|_2 &= \left(\Delta x\sum_i |g_i^n|^2\right)^{\frac{1}{2}} \\
&\leq \Delta x^{\frac{1}{2}}\frac{|\omega_i|}{p!}\left(\frac{1}{p+1}\right)^{\frac{1}{2}}\left(\frac{1}{p\Delta x}\right)^{\frac{1}{2}}\left(\sum_i\left\|u^{(p+1)}\right\|_{L^2(I_i)}^2\right)^{\frac{1}{2}}.
\end{aligned}$$

We notice that

$$\sum_i\left\|u^{(p+1)}\right\|_{L^2(I_i)}^2 = p\left\|u^{(p+1)}\right\|_{L^2(\mathbb{R})}^2 = p\left\|u_0^{(p+1)}\right\|_{L^2(\mathbb{R})}^2.$$

Finally we use all these inequalities and obtain the claim (38) after simplification. □

Proposition 10. *Assume that $p \in \{2k, 2k+1, 2k+2\}$. Then the following error estimate hold for all n such that $n\Delta t \leq T$:*

$$\|e^n\|_2 \leq C_{k,p}\,(1-\nu)\,\frac{v\Delta s\Delta x^p}{\Delta t}T\|u_0^{(p+1)}\|_{L^2(\mathbb{R})}. \tag{40}$$

Proof. One gets from (33-34) $\left\|e^{n+1}\right\|_2 \leq \|e^n\|_2 + \|g^n\|_2$, that is, since $e^0 = 0$,

$$\|e^n\|_2 \leq nC_{k,p}\nu(1-\nu)\left\|u_0^{(p+1)}\right\|_{L^2(\mathbb{R})}\Delta x^{p+1}.$$

It proves the claim after rearrangements. □

Remark 11. *A crude estimate based on the Stirling formula shows that*

$$C_{k,p} = O\left(2^{-2k}\right)$$

is very small for large p.

Proof. Consider for example the case $p = 2k + 1$: the constant is

$$C_{k,p} = \frac{1}{2^{2k+2}} \times \frac{2^{2k+1}}{(p+1)^{\frac{1}{2}} \begin{pmatrix} 2k+2 \\ k+1 \end{pmatrix}}. \tag{41}$$

Since the Stirling formula shows that $\frac{2^{2k+1}}{\begin{pmatrix} 2k+2 \\ k+1 \end{pmatrix}} \approx Ck^{\frac{1}{2}}$ for some constant C when k is large, it shows the bound of the remark. □

4.4 Truncation-based error estimate in L^2

In this section we propose another analysis of the error, which is based on the analysis of the truncation error. The philosophy is rather different. The analysis will fill the gap between interpolation error analysis based on (35) and the Fourier analysis as used in the work of Thomée [16]. The conclusions will be the same, but with a slightly better constant for large p. This analysis is confirmed with a different approach discussed in the appendix. In both cases the key is the use of advanced formulas which ultimately may be used to provide sharper L^2 estimates for B-splines.

4.4.1 Truncation error

It is convenient to write the scheme in the incremental form

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + v \frac{u_{j+\frac{1}{2}}^n - u_{j-\frac{1}{2}}^n}{\Delta x} = 0 \tag{42}$$

where $u_{j+\frac{1}{2}}^n$ is the flux. The Fourier symbol of the flux is the function $\psi \mapsto \gamma_{\nu,k,p}(\psi)$ such that

$$\bar{u}_{j+\frac{1}{2}}(\psi) = \gamma_{\nu,k,p}(\psi)\bar{u}_j(\psi),$$

where $\bar{u}$ is defined by (27). We deduce from the expressions (28) and (42) the following expression for $\gamma_{\nu,k,p}$:

Proposition 12. *The Fourier symbol of the flux is*

$$\gamma_{\nu,k,p}(\psi) = \frac{\lambda_{\nu,k,p}(\psi) - 1}{(e^{\mathbf{i}\psi} - 1)} \frac{v\Delta t}{\Delta x}. \tag{43}$$

The method we propose here is use this expression of the Fourier symbol of the flux in the truncation error ε_i^n to obtain the analogous of Proposition 2 for arbitrary order schemes. The truncation error can here be written under the form

$$\varepsilon_i^n = \frac{u((n+1)\Delta t, x_i) - u(n\Delta t, x_i)}{\Delta t} + v \frac{\phi(u(n\Delta t, \cdot))(x_i) - \phi(u(n\Delta t, \cdot))(x_{i-1})}{\Delta x}. \tag{44}$$

Let

$$\widehat{u}(\cdot,\theta)=\int_{\mathbb{R}}u(\cdot,x)\,e^{-i\theta x}dx. \tag{45}$$

be the Fourier transform $\widehat{u}$ in the x variable of u. Then the Fourier symbol of the operator ϕ in (44) is given by $\gamma_{\nu,k,p}\left(\theta\Delta x\right)$.

Moreover, we easily verify, thanks to the relations $u(n\Delta t+\Delta t,x)=u(n\Delta t,x-v\Delta t)$ and $v\Delta t=v\Delta s-(r+1)\Delta x$ that

$$\widehat{u}(n\Delta t+\Delta t,\theta)=e^{i\theta((r+1)-\nu)\Delta x}\widehat{u}(n\Delta t,\theta).$$

Assuming that $u(t,\cdot)\in L^2(\mathbb{R})$ for all t, we have the following relation

$$u(\cdot,x)=\frac{1}{2\pi}\int_{\mathbb{R}}\widehat{u}(\cdot,\theta)e^{\mathbf{i}\theta x}d\theta. \tag{46}$$

Now let express with thanks to the inversion formulae (46) the truncation error :

$$\begin{aligned}\varepsilon_i^n&=\frac{1}{2\pi}\int_{\mathbb{R}}\left(\frac{e^{\mathbf{i}((r+1)-\nu)\theta\Delta x}-1}{\Delta t}+v\frac{\gamma_{\nu,k,p}\left(\theta\Delta x\right)\left(1-e^{\mathbf{i}\theta\Delta x}\right)}{\Delta x}\right)\widehat{u}(n\Delta t,\theta)e^{\mathbf{i}\theta x_i}d\theta\\&=\frac{1}{2\pi}\int_{\mathbb{R}}\left(\frac{e^{\mathbf{i}((r+1)-\nu)\theta\Delta x}-\lambda_{\nu,k,p}\left(\theta\Delta x\right)}{\Delta t}\right)\widehat{u}(n\Delta t,\theta)e^{\mathbf{i}\theta x_i}d\theta\\&=\frac{1}{2\pi}\frac{\Delta s}{\Delta t}\int_{\mathbb{R}}\left(\frac{e^{-\mathbf{i}\nu\theta\Delta x}-\mu_{\nu,k,p}\left(\theta\Delta x\right)}{\Delta s}\right)\widehat{u}(n\Delta t,\theta)e^{\mathbf{i}\theta x_{i+r+1}}d\theta\end{aligned} \tag{47}$$

thanks to (43) and (31). We now recognize at the right side the truncation operator $\overline{\varepsilon}_{i+r+1}^n$ of the semi-lagrangian scheme but with a time step equal to Δs. At last, we obtain the following Proposition,

Proposition 13. *One has the formula $\varepsilon_i^n=\frac{\Delta s}{\Delta t}\overline{\varepsilon}_{i+r+1}^n$ for all i and n, and for all coefficients ν,p,k.*

This expression is the generalization of Proposition 2 at any order.

4.4.2 Bounds

To continue we use a formula which is proved in [5]. We consider here the scheme of time step Δs and the amplification factor $\mu_{\nu,k,p}$; the truncation error corresponding to this scheme is $\bar{\varepsilon}_i^n$. According to [5] (1.5), the amplification factor can be written on the following form :

$$\mu_{\nu,k,p}\left(\psi\right)=e^{-\mathbf{i}\nu\psi}(1-\alpha(\psi))$$

where we have set for convenience

$$\alpha(\psi)=\mathbf{i}^{p+1}\alpha_{k,p}(\nu)2^p\int_0^{\psi}\sin^p\left(\frac{\varphi}{2}\right)e^{\mathbf{i}(k-\frac{p}{2}+\nu)(\varphi)}d\varphi \tag{48}$$

and

$$\alpha_{k,p}(\nu) = \frac{\prod_{q=0}^{p}(k+\nu-q)}{p!}. \tag{49}$$

Therefore the corresponding truncation error is, thanks to (47)

$$\overline{\varepsilon}_i^n = \frac{1}{2\pi}\frac{1}{\Delta s}\int_{\mathbb{R}} \alpha(\theta\Delta x)\widehat{u}(n\Delta t,\theta)e^{\mathbf{i}\theta x_i}d\theta \tag{50}$$

A direct use of the kernel α is possible using the theory of Thomée [15, 16]. But it is much more efficient for our purposes to use the following trick. We rewrite first (50) with a backward Fourier transform

$$\widehat{u}(n\Delta t,\theta) = \int_{\mathbb{R}} u(n\Delta t,x)e^{-\mathbf{i}\theta x}dx.$$

One obtains

$$\overline{\varepsilon}_i^n = \frac{1}{2\pi\Delta s}\int_{\mathbb{R}} u(n\Delta t,x)\left(\int_{\mathbb{R}} \alpha(\theta\Delta x)e^{\mathbf{i}\theta(x_i-x)}d\theta\right)dx. \tag{51}$$

Then we integrate $p+1$ times by parts with respect to x

$$\overline{\varepsilon}_i^n = \frac{1}{2\pi\Delta s}\int_{\mathbb{R}} \partial_x^{(p+1)}u(n\Delta t,x)\left(\int_{\mathbb{R}} \frac{\alpha(\theta\Delta x)}{\mathbf{i}^{p+1}\theta^{p+1}}e^{\mathbf{i}\theta(x_i-x)}d\theta\right)dx.$$

A restriction is that we must consider $p \geq 2$ so that the internal integral is absolutely convergent. This is not a real restriction since the case $p = 1$ is elementary and can be treated separately as in Section 3. The kernel in the integral is

$$\beta_i(x) := \int_{\mathbb{R}} \frac{\alpha(\theta\Delta x)}{\mathbf{i}^{p+1}\theta^{p+1}}e^{\mathbf{i}\theta(x_i-x)}d\theta. \tag{52}$$

One has the property :

Lemma 14. *The kernel β_i has compact support in $[x_i+(k-p)\Delta x, x_i+k\Delta x]$.*

Proof. The formula $\overline{\varepsilon}_i^n = \frac{1}{2\pi\Delta s}\int_{\mathbb{R}} \partial_x^{(p+1)}u(n\Delta t,x)\beta_i(x)dx$ shows that the truncation error is a function of $\partial_x^{(p+1)}u$ only, but not of the other derivatives. It is evident that there exists an integral representation

$$\overline{\varepsilon}_i^n = \int_{x_i+(k-p)\Delta x}^{x_i+k\Delta x} K(x)\partial_x^{(p+1)}u(n\Delta t,x)dx$$

in the compact interval $[x_i+(k-p)\Delta x, x_i+k\Delta x]$ for some kernel K. This is due to two facts: the stencil of the scheme is compact; the order of the scheme is p. Since the function $\partial_x^{(p+1)}u$ is arbitrary in these representation formulas, it implies that

$$\begin{aligned}\beta_i(x) &= 2\pi\Delta s\, K(x) \quad \text{if } x_i+(k-p)\Delta x < x < x_i+k\Delta x,\\ &= 0 \quad \text{otherwise.}\end{aligned}$$

This is why β_i has indeed a compact support

$$\overline{\varepsilon}_i^n = \frac{1}{2\pi\Delta s}\int_{x_i+(k-p)\Delta x}^{x_i+k\Delta x} \partial_x^{(p+1)} u(n\Delta t, x)\beta_i(x)dx. \tag{53}$$

See also Remark 17. □

Proposition 15. *There exists $C > 0$ such that*

$$\|\overline{\varepsilon}^n\|_2 \leq C\frac{C_{k,p}}{p^{\frac{1}{4}}}(1-\nu)(v\Delta x^p)\left\|u_0^{(p+1)}\right\|_{L^2(\mathbb{R})} \tag{54}$$

where $C_{k,p}$ is defined by (39).

Proof. Thanks to (53), we get

$$|\overline{\varepsilon}_i^n| \leq \frac{1}{2\pi\Delta s}\|\beta_i\|_{L^2(\mathbb{R})}\left\|\partial_x^{(p+1)}u(n\Delta t,\cdot)\right\|_{L^2(x_i+(k-p)\Delta x, x_i+k\Delta x)}.$$

Since

$$\sum_i \left\|\partial_x^{(p+1)}u(n\Delta t,\cdot)\right\|^2_{L^2(x_i+(k-p)\Delta x, x_i+k\Delta x)} = p\left\|\partial_x^{(p+1)}u(n\Delta t,\cdot)\right\|^2_{L^2(\mathbb{R})}$$

one gets that

$$\|\overline{\varepsilon}^n\|_2 = \left(\Delta x\sum_i |\overline{\varepsilon}_i^n|^2\right)^{\frac{1}{2}} \leq \frac{(p\Delta x)^{\frac{1}{2}}}{2\pi\Delta s}\sup_i\|\beta_i\|_{L^2(\mathbb{R})}\left\|\partial_x^{(p+1)}u(n\Delta t,\cdot)\right\|_{L^2(\mathbb{R})}.$$

Now let us estimate $\|\beta_i\|_{L^2(\mathbb{R})}$, for $i \in \mathbb{Z}$. According to (52) and Proposition 14, we have

$$\begin{aligned}
\|\beta_i\|_{L^2(\mathbb{R})} &= \left(\int_{\mathbb{R}}\left|\int_{\mathbb{R}}\frac{\alpha(\theta\Delta x)}{\mathbf{i}^{p+1}\theta^{p+1}}e^{\mathbf{i}\theta(x_i-x)}d\theta\right|^2 dx\right)^{1/2} \\
&\leq C\left(\int_{\mathbb{R}}\frac{|\alpha(\theta\Delta x)|^2}{\theta^{2p+2}}d\theta\right)^{\frac{1}{2}} \\
&\leq C|\alpha_{k,p}(\nu)|2^p\left(\int_{\mathbb{R}}\left(\int_0^{\theta\Delta x}|\sin^p(\frac{\varphi}{2})|d\varphi\right)^2\frac{d\theta}{\theta^{2p+2}}\right)^{\frac{1}{2}}.
\end{aligned} \tag{55}$$

thanks to (48). Elementary changes of variable show that

$$\int_{\mathbb{R}}\left(\int_0^{\theta\Delta x}|\sin^p(\frac{\varphi}{2})|d\varphi\right)^2\frac{d\theta}{\theta^{2p+2}} = \frac{\Delta x^{2p+1}}{2^{2p-1}}\int_0^\infty\left(\int_0^\theta|\sin^p(\varphi)|d\varphi\right)^2\frac{d\theta}{\theta^{2p+2}}.$$

Using (82) (see Annexe B) we get that

$$\|\beta_i\|_{L^2(\mathbb{R})} \leq M \frac{\Delta x^{p+\frac{1}{2}} |\alpha_{k,p}(\nu)|}{p^{1+\frac{1}{4}}}$$

from which we deduce

$$\|\bar{\varepsilon}^n\|_2 \leq \frac{C}{\Delta s} \frac{|\alpha_{k,p}(\nu)|}{p^{\frac{1}{2}+\frac{1}{4}}} \Delta x^{p+1} \left\| u_0^{(p+1)} \right\|_{L^2(\mathbb{R})}.$$

Finally we notice that $|\alpha_{k,p}(\nu)| = \frac{1}{p!} \left| \prod_{l=k-p}^{k} (l+\nu) \right|$ and thanks to estimate (39), we obtain

$$\|\bar{\varepsilon}^n\|_2 \leq C \frac{C_{k,p}}{p^{\frac{1}{4}}} (1-\nu)\nu \frac{\Delta x^{p+1}}{\Delta s} \left\| u_0^{(p+1)} \right\|_{L^2(\mathbb{R})}. \tag{56}$$

Since $\nu = v\Delta s/\Delta x$, the claim is proved. □

Theorem 16. *Assume $p = 2k$, $2k+1$ or $2k+2$. One has the optimal error estimate for all n such that $n\Delta t \leq T$:*

$$\|e^n\|_2 \leq D_{k,p} \left(1 - \frac{v\Delta s}{\Delta x}\right) \frac{v\Delta s \Delta x^p}{\Delta t} T \|u_0^{(p+1)}\|_{L^2(\mathbb{R})} \tag{57}$$

Proof. From Proposition 13 one gets the estimate for the true truncation error

$$\|\varepsilon^n\|_2 \leq D_{k,p}(1-\nu) \frac{v\Delta s \Delta x^p}{\Delta t} \left\| u_0^{(p+1)} \right\|_{L^2(\mathbb{R})}.$$

Since by definition of the truncation error

$$e^{n+1} = \mathcal{R}_{\nu,k,p}\ e^n + \Delta t \varepsilon^n \tag{58}$$

one has the estimate $\left\|e^{n+1}\right\|_2 \leq \|e^n\|_2 + \Delta t \|\varepsilon^n\|_2$. After summation over n time steps ($n\Delta t \leq T$), one gets the result. □

Remark 17. *The comparison of (34) and (58) shows that $g^n = \Delta t \varepsilon^n$, that is the interpolation error is proportional to the truncation error. In consequence the kernel Q_i^{p+1} is proportional to the kernel β_i. So we understand that the Fourier formula (55) is just a more accurate way to bound the L^2 norm of the B-spline, at least more accurate than the crude estimate of $\left\|Q_i^{p+1}\right\|_{L^\infty}$ together with*

$$\left\|Q_i^{p+1}\right\|_{L^2} \leq \left\|Q_i^{p+1}\right\|_{L^1}^{\frac{1}{2}} \left\|Q_i^{p+1}\right\|_{L^\infty}^{\frac{1}{2}} \tag{59}$$

that was used in the proof of Proposition 9. This is why the constant $D_{k,p}$ is better than the constant $C_{k,p}$ by a factor $p^{-\frac{1}{4}}$ for large p. More material about the L^∞ norm of Q_i^{p+1} is provided in Appendix A.

4.5 L^∞ and L^1 error estimates for odd order schemes

We first have the fundamental result [5].

Proposition 18. *Assume that $p = 2k+1$. Then powers of the iteration operator are uniformly bounded in L^q norms. More precisely there exists $C > 0$ such that*

$$\left\| \left(\mathcal{R}_{\nu,k,p}\right)^l \right\|_q \leq C, \quad \forall l \in \mathbb{N},\ \forall\ 1 \leq q \leq \infty,\ \forall v \frac{\Delta t}{\Delta x} \leq 1. \tag{60}$$

It must be understood that $r = -1$.

The shifting technique of semi-lagrangian schemes allows to treat all values of the CFL number, that is all values of r. We obtain

Proposition 19. *Assume that $p = 2k+1$. Then powers of the iteration operator are uniformly bounded in L^q norms. More precisely there exists $C > 0$ such that*

$$\left\| \left(\mathcal{R}_{\nu,k,p}\right)^l \right\|_q \leq C, \quad \forall l \in \mathbb{N},\ \forall\ 1 \leq q \leq \infty,\ \forall r. \tag{61}$$

It is then easy to generalize the approach developed in Proposition 9 to obtain

Theorem 20. *Assume $p = 2k + 1$. One has the optimal error estimate for all n such that $n\Delta t \leq T$*

$$\|e^n\|_q \leq E_{k,p} \left(1 - \frac{v\Delta s}{\Delta x}\right) \frac{v\Delta s \Delta x^p}{\Delta t} \|u_0^{(p+1)}\|_{L^q(\mathbb{R})}. \tag{62}$$

5 Periodic domains

All results of the previous sections hold true in periodic domains. In particular we will use the L^2 error estimate (40) or (57) in the next section.

6 The Vlasov-Poisson equation

We propose in this section to apply estimates obtained in the previous section to improve the error estimate proposed in [10] in the context of semi-lagrangian schemes for the Vlasov-Poisson system. The adimensional Vlasov-Poisson equation in one dimension in space and in velocity reads

$$\frac{\partial f}{\partial v} + v\frac{\partial f}{\partial x} + E(t,x)\frac{\partial f}{\partial v} = 0, \tag{63}$$

with

$$\frac{\partial E}{\partial x}(t,x) = \int_{-\infty}^{+\infty} f(t,x,v)dv - 1, \tag{64}$$

where $f(t,x,v)$ is the distribution function of charged particles (ions or electrons) and $E(t,x)$ the self-consistent electric field. We consider periodic boundary conditions on the variable $x \in [0,L]$, that is

$$f(t,0,v) = f(t,L,v), \quad v \in \mathbb{R}, \quad t \geq 0, \tag{65}$$

and

$$E(t,0) = E(t,L). \tag{66}$$

In order to have a well-posed problem, we add to equation (63)–(66) the zero-mean condition on the field E

$$\int_0^L E(t,x)dx = 0, \quad t \geq 0 \tag{67}$$

and a initial condition

$$f(0,x,v) = f_0(x,v), \quad x \in L, \quad v \in \mathbb{R}. \tag{68}$$

The electric field E is given by

$$E(t,x) = \int_0^L K(x,y) \left(\int_{\mathbb{R}} f(t,y,v)dv - 1 \right) dy$$

where K is computed from the kernel of the Poisson equation.

We refer to [12], for example, for the existence, uniqueness and regularity of the Vlasov-Poisson system (63)–(68). We will also assume that the initial data has compact support. In consequence the solution has also a compact support in velocity. It is enough to consider the domain $v \in [-v_{\max}, v_{\max}]$ for $t \leq T$. The domain of simulation can be taken as

$$\Omega = [0,L]_{\text{per}} \times [-v_{\max}, v_{\max}]_{\text{per}}.$$

6.1 Numerical strategy and convergence

Let $N_x, N_v \in \mathbb{N}^*$. We a define numerical approximation $f^{(n)} \in \mathbb{R}^{N_x \times N_v}$, by

$$f_{i,j}^{(n)}, \ i = 0, \ldots, N_x - 1, \ j = 0, \ldots, N_v - 1, \ n \in \mathbb{N}.$$

The L^2 norm of a discrete function is defined in (5).
We introduce the reconstruction operator in the x direction $\mathcal{R}_x : \mathbb{R}^{N_x \times N_v} \to L^\infty([0,L]) \times \mathbb{R}^{N_v}$ or $\mathcal{R}_x : \mathbb{R}^{N_x} \to L^\infty([0,L])$

$$f(\cdot) = \mathcal{R}_x f.$$

In practice we use a Lagrangian interpolation of order $p+1$ in the x direction as described in (24).
We have the same definition for the reconstruction in the v direction $\mathcal{R}_v : \mathbb{R}^{N_x \times N_v} \to \mathbb{R}^{N_x} \times L^\infty([-v_{\max}, v_{\max}])$ or $\mathcal{R}_v : \mathbb{R}^{N_v} \to L^\infty([-v_{\max}, v_{\max}])$

$$f(\cdot) = \mathcal{R}_v f,$$

with also a Lagrangian reconstruction of order $p+1$ but in the v direction. We also define the reconstruction in both directions $\mathcal{R}_x \otimes \mathcal{R}_v : \mathbb{R}^{N_x \times N_v} \to L^\infty(\Omega_m)$ which is the "tensorial" product of the one dimensional reconstructions $\mathcal{R}_x$ and $\mathcal{R}_v$: that is we reconstruct first in the v direction and second in the x direction. Notice that the result may be different if one reconstructs first in the x direction and after that in the v direction, unless a compatibility condition is satisfied for commutativity. In the context of this work, such a compatibility condition is not required: the only thing is to guarantee the error estimate (72). For a continuous function $g \in \mathcal{C}(\Omega)$, we define the projected discrete function $\Pi g \in \mathbb{R}^{N_x \times N_v}$, by

$$(\Pi g)_{i,j} = g(x_i, v_j),\ i = 0, \ldots, N_x - 1,\ j = 0, \ldots, N_v - 1.$$

Definition 21. *We define the discrete transport operator in the x direction based on a semi-lagrangian scheme of order $p+1$ by $\mathcal{T}_{x,s} : \mathbb{R}^{N_x \times N_v} \to \mathbb{R}^{N_x \times N_v}$.*

Definition 22. *We define the discrete transport operator in the v direction based on a semi-lagrangian scheme of order $p+1$ by $\mathcal{T}_{v,E,s} : \mathbb{R}^{N_x \times N_v} \to \mathbb{R}^{N_x \times N_v}$.*

A fundamental result deduced from the stability property (33) is

$$\|\mathcal{T}_{x,s}\|_2 \leq 1 \text{ and } \|\mathcal{T}_{v,E,s}\|_2 \leq 1 \tag{69}$$

for all s and E. This is automatically true is one uses the semi-Lagrangian schemes or Strang's stencil discussed in the first part of this paper.

We will also use the exact transport operators $\widetilde{\mathcal{T}_{x,s}}$ and $\widetilde{\mathcal{T}_{v,E,s}}$ defined for functions.

6.1.1 Algorithm

The Vlasov-Poisson discrete scheme [6, 1, 10] reads

$$f^{(n+1)} = \mathcal{T}_{x,\Delta t/2}\mathcal{T}_{v,E_h^n,\Delta t}\mathcal{T}_{x,\Delta t/2}f^{(n)},\ n \in \mathbb{N},\ f^{(0)} = \Pi f(0),$$

with the electric field calculated with the following approximation of the exact kernel

$$E_h^n(x) = \int_0^L K(x,y)\left(\int_{-v_{\max}}^{v_{\max}} \left(\mathcal{R}_x \otimes \mathcal{R}_v \mathcal{T}_{x,\Delta t/2} f^{(n)}\right)(y,v)dv - 1\right)dy. \tag{70}$$

Remark 23. *It can be checked that (70) is equivalent to*

$$E_h^n(x) = \int_0^L K(x,y)\mathcal{R}_x\left(\int_{-v_{\max}}^{v_{\max}} \left(\mathcal{R}_v \mathcal{T}_{x,\Delta t/2} f^{(n)}\right)(\cdot,v)dv - 1\right)dy$$

which is more convenient for implementation purposes. On can notice that the internal integral is

$$\rho_i = \int_{-v_{\max}}^{v_{\max}} \mathcal{R}_v\left(\mathcal{T}_{x,\Delta t/2} f_{i,\cdot}^{(n)}\right)(v)dv - 1 = \Delta v \sum_{j=0}^{N_v-1} \mathcal{T}_{x,\Delta t/2} f_{i,j}^{(n)} - 1,$$

which provides an easy way to compute it. Another possibility could be to define the discrete electric field with

$$E_h^n(x) = \int_0^L K(x,y) \left(\int_{-v_{\max}}^{v_{\max}} \widetilde{\mathcal{T}_{x,\Delta t/2}} \mathcal{R}_x \otimes \mathcal{R}_v f^{(n)}(y,v) dv - 1 \right) dy.$$

This formula seems less evident to implement. For theoretical considerations all formulas are equivalent provided the high order error formula (72) holds.

One notices that E_h^n may be calculated for any x in the domain, not only at grid points $x = x_i$. We will use this trick in the term ε_4 in the decomposition that follows.

6.1.2 Error decomposition

The numerical error is by definition $e^{(n)} \in \mathbb{R}^{N_x \times N_v}$

$$e^{(n)} = \Pi f(t_n) - f^{(n)}.$$

We have the following error decomposition

$$e^{(n+1)} = \varepsilon_1 + \varepsilon_2 + \mathcal{T}_{x,\Delta t/2}\varepsilon_3 + \mathcal{T}_{x,\Delta t/2}\varepsilon_4 + \mathcal{T}_{x,\Delta t/2}\mathcal{T}_{v,E_h^n,\Delta t}\varepsilon_5,$$

with

$$\begin{aligned}
\varepsilon_1 &= \Pi f(t_{n+1}) - \Pi \widetilde{\mathcal{T}_{x,\Delta t/2}} \widetilde{\mathcal{T}_{v,E,\Delta t}} \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n), \\
\varepsilon_2 &= \left(\Pi \widetilde{\mathcal{T}_{x,\Delta t/2}} - \mathcal{T}_{x,\Delta t/2}\Pi \right) \widetilde{\mathcal{T}_{v,E,\Delta t}} \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n). \\
\varepsilon_3 &= \Pi \left(\widetilde{\mathcal{T}_{v,E,\Delta t}} - \widetilde{\mathcal{T}_{v,E_h^n,\Delta t}} \right) \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n), \\
\varepsilon_4 &= \Pi \widetilde{\mathcal{T}_{v,E_h^n,\Delta t}} \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n) - \mathcal{T}_{v,E_h^n,\Delta t} \Pi \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n), \\
\varepsilon_5 &= \Pi \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n) - \mathcal{T}_{x,\Delta t/2} f^{(n)},
\end{aligned}$$

6.1.3 Time error of the Strang's splitting

We recall the following result (cf [1] for example)

Lemma 24. *If $f(t_n)$ is bounded in $W^{1,\infty}(\Omega)$, then*

$$\| f(t_{n+1}) - \widetilde{\mathcal{T}_{x,\Delta t/2}} \widetilde{\mathcal{T}_{v,E,\Delta t}} \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n) \|_{L^\infty(\Omega)} \leq C_T \Delta t^3.$$

We then get

$$\|\varepsilon_1\|_2 \leq C_1 \Delta t^3, \tag{71}$$

since the domain is bounded and with periodic conditions, and the "initial data" $f(t_n)$ has compact support inside Ω.

6.1.4 Estimate for the electric field

The third term in the decomposition is

$$\begin{aligned}(\varepsilon_3)_{i,j} =& f\left(t_n, x_i - (v_j - E^n(x_i)\Delta t)\Delta t/2, v_j - E^n(x_i)\Delta t\right)\\ &- f\left(t_n, x_i - (v_j - E_h^n(x_i)\Delta t)\Delta t/2, v_j - E_h^n(x_i)\Delta t\right),\end{aligned}$$

where we denote E^n the electric field calculated from $\widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n)$

$$E^n(x) := \int_0^L K(x,y)\left(\int_{\mathbb{R}}\left(\widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n)\right)(y,v)dv - 1\right)dy$$

One has

$$E^n(x_i) - E_h^n(x_i) = \int_0^L K(x_i,y)\int_{-v_{\max}}^{v_{\max}} g(y,v)dvdy,$$

where

$$\begin{aligned}g =& \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n) - \mathcal{R}_x \otimes \mathcal{R}_v \Pi \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n)\\ &+ \mathcal{R}_x \otimes \mathcal{R}_v (\Pi \widetilde{\mathcal{T}_{x,\Delta t/2}} - \mathcal{T}_{x,\Delta t/2}\Pi) f(t_n)\\ &+ \mathcal{R}_x \otimes \mathcal{R}_v \mathcal{T}_{x,\Delta t/2}(\Pi f(t_n) - f^{(n)}).\end{aligned}$$

By using (13) in [10], we get the following a priori estimate

$$\max_{i\in\{0,\dots,N_x-1\}} |E^n(x_i) - E_h^n(x_i)| \le C\left(\max(\Delta x, \Delta v)^{p+1} + \|e^{(n)}\|_2\right). \tag{72}$$

Moreover, since $E^n \in \mathrm{L}^\infty(O,L)$, we get

$$\begin{aligned}\max_{i\in\{0,\dots,N_x-1\}} |E_h^n(x_i)| &\le C'(1 + \|e^{(n)}\|_2 + \max(\Delta x, \Delta v)^{p+1})\\ &\le C''(1 + \|f^{(n)}\|_2 + \|\Pi f(t_n)\|_2)\end{aligned} \tag{73}$$

and then

$$\sup_{n\le \frac{T}{\Delta t}} \max_{i\in\{0,\dots,N_x-1\}} |E_h^n(x_i)| < +\infty \tag{74}$$

by using the fact that

$$\|\Pi f(t_n)\|_2 \le 2v_{\max} L \|f(t_n)\|_{L^\infty(\Omega)} \le 2v_{\max} L \|f(0)\|_{L^\infty(\Omega)},$$

and, from the stability estimates (69)

$$\|f^{(n)}\|_2 \le \|f^{(0)}\|_2 \le 2v_{\max} L \|f(0)\|_{L^\infty(\Omega)}.$$

6.1.5 Final proof of Theorem 1

Proof. The triangular inequality implies that

$$\|e^{(n+1)}\|_2 \le \|\varepsilon_1\|_2 + \|\varepsilon_2\|_2 + \|\varepsilon_3\|_2 + \|\varepsilon_4\|_2 + \|\varepsilon_5\|_2.$$

The first term is bounded using (71). The second term is bounded as

$$\|\varepsilon_2\|_2 \leq C_2 \nu \Delta x^{p+1}$$

using (34)-(38). Using (72) the third term is bounded as

$$\|\varepsilon_3\|_2 \leq C_3 \Delta t \left(\max(\Delta x, \Delta v)^{p+1} + \|e^{(n)}\|_2 \right).$$

The fourth term can also be written

$$\varepsilon_4 = \left(\Pi \widetilde{\mathcal{T}_{v,E_h^n,\Delta t}} - \mathcal{T}_{v,E_h^n,\Delta t} \Pi \right) \widetilde{\mathcal{T}_{x,\Delta t/2}} f(t_n) \tag{75}$$

and then

$$\|(\varepsilon_4)\|_2 \leq C_4 \tau_v \Delta v^{p+1}$$

where τ_v is an upper bound of all $\tilde{\nu} = \frac{|E_h^n(x_i)|\Delta s}{\Delta v}$ for all possible columns in the domain of computation. One must be careful that Δs is a function of v, but anyway we have $\Delta s \leq \Delta t$. Therefore, a sharp estimate for τ_v is, thanks to (74),

$$\tau_v = \min \left(\sup_{n \leq \frac{T}{\Delta t}} \max_{i \in \{0,\ldots,N_x-1\}} |E_h^n(x_i)| \frac{\Delta t}{\Delta v}, 1 \right). \tag{76}$$

Finally the fifth term can also be written

$$\varepsilon_5 = \left(\Pi \widetilde{\mathcal{T}_{x,\Delta t/2}} - \mathcal{T}_{x,\Delta t/2} \Pi \right) f(t_n) + \mathcal{T}_{x,\Delta t/2} e^{(n)}, \tag{77}$$

and then is bounded as

$$\|\varepsilon_5\|_2 \leq \| \left(\Pi \widetilde{\mathcal{T}_{x,\Delta t/2}} - \mathcal{T}_{x,\Delta t/2} \Pi \right) f(t_n)\|_2 + \|e^{(n)}\|_2$$

which, using (34)-(38), yields

$$\|\varepsilon_5\|_2 \leq C_5 \tau_x \Delta x^{p+1} + \|e^{(n)}\|_2.$$

In this formula τ_x is an upper bound of all $\nu = \frac{v\Delta s}{\Delta x}$ for all possible lines in the domain of computation. Since $v\Delta s \leq \Delta x$ and $\Delta s \leq \Delta t$, one can take as sharp estimate for τ_x

$$\tau_x = \min \left(v_{\max} \frac{\Delta t}{\Delta x}, 1 \right). \tag{78}$$

Assuming for simplicity that $\Delta v = \alpha \Delta x$ with a parameter α which is independent of Δx, one finds that, since $\Delta t \leq T,\ \Delta x \leq L$,

$$\begin{aligned} \Delta t \max(\Delta x, \Delta v)^{p+1} &\leq \max(\alpha^{p+1}, 1) \Delta t \Delta x^{p+1} \\ &\leq \max(\alpha^{p+1}, 1)(L+T) \min(\frac{\Delta t}{\Delta x}, 1) \Delta x^{p+1}, \end{aligned}$$

and thus

$$\|e^{(n+1)}\|_2 \leq \|e^{(n)}\|_2 + c\left(\min\left(\frac{\Delta t}{\Delta x}, 1\right)\Delta x^{p+1} + \Delta t\|e^{(n)}\|_2 + \Delta t^3\right)$$
$$\leq e^{c\Delta t}\|e^{(n)}\|_2 + c\left(\min\left(\frac{\Delta t}{\Delta x}, 1\right)\Delta x^{p+1} + \Delta t^3\right)$$

since $1 + c\Delta t \leq e^{c\Delta t}$. Therefore after summation

$$\|e^{(n)}\|_2 \leq e^{cn\Delta t}\|e^{(0)}\|_2 + c\left(\min\left(\frac{\Delta t}{\Delta x}, 1\right)\Delta x^{p+1} + \Delta t^3\right)\left(1 + e^{c\Delta t} + e^{c2\Delta t} + \cdots + e^{c(n-1)\Delta t}\right).$$

Since $e^{(0)} = 0$ and $n\Delta t \leq T$, it yields

$$\|e^{(n)}\|_2 \leq \left(\min\left(\frac{\Delta t}{\Delta x}, 1\right)\Delta x^{p+1} + \Delta t^3\right)\frac{C}{\Delta t}, \quad C > 0.$$

It ends the proof of Theorem 1. □

A More about the L^∞ norm of the B-splines

We give more material about B-splines and explain how to recover an optimal L^∞ norm starting from advanced results discussed in [11]. It provides an alternative way to analyze (59).

A.1 B-Splines

We consider here the B-spline over the interval $[0, 1]$ as in [11]. In the even case, we consider the $B_{2d+1,\alpha}$ spline over the $2d + 2$ points

$$0 < \frac{1}{2d} < \cdots < \frac{d}{2d} \leq \frac{d+\alpha}{2d} < \frac{d+1}{2d} < \cdots < \frac{2d-1}{2d} < 1,$$

and in the odd case, we consider the $B_{2d+2,\alpha}$ spline over the $2d + 3$ points

$$0 < \frac{1}{2d+1} < \cdots < \frac{d}{2d+1} \leq \frac{d+\alpha}{2d+1} < \frac{d+1}{2d+1} < \cdots < \frac{2d}{2d+1} < 1.$$

The B-splines are here defined so that

$$\int_0^1 B_{2d+1,\alpha}(x)dx = \frac{1}{2d+1}, \ \int_0^1 B_{2d+2,\alpha}(x)dx = \frac{1}{2d+2}.$$

We then look for $\sup_{0\leq\alpha<1}\|B_{2d+1,\alpha}\|_{L^\infty}$ and $\sup_{0\leq\alpha<1}\|B_{2d+2,\alpha}\|_{L^\infty}$. We have (see Theorem 5, in [11])

$$B_{2d+1,\alpha}(x) = \frac{x - \frac{d+\alpha}{2d}}{2d}B'_{2d+1,\alpha}(x) + \tilde{B}_{2d}(x),$$

where $\tilde{B}_{2d}$ is the uniform B-spline with $2d+1$ points, and also

$$B_{2d+2,\alpha}(x) = \frac{x - \frac{d+\alpha}{2d+1}}{2d+1} B'_{2d+2,\alpha}(x) + \tilde{B}_{2d+1}(x),$$

where $\tilde{B}_{2d+1}$ is the uniform B-spline with $2d+2$ points. Now, let $x^*_{2d+1,\alpha}$ such that $B_{2d+1,\alpha}(x^*_{2d+1,\alpha}) = \|B_{2d+1,\alpha}\|_{L^\infty}$. We have $B'_{2d+1,\alpha}(x^*_{2d+1,\alpha}) = 0$, and thus $\|B_{2d+1,\alpha}\|_\infty = \tilde{B}_{2d}(x^*_{2d+1,\alpha})$. We also have $B_{2d+1,0}(1/2) = \tilde{B}_{2d}(1/2)$. We then get

$$B_{2d+1,0}(1/2) = \tilde{B}_{2d}(1/2) = \|\tilde{B}_{2d}\|_{L\infty} \geq \tilde{B}_{2d}(x^*_{2d+1,\alpha}) = \|B_{2d+1,\alpha}\|_{L^\infty},$$

which means that (see proof of Theorem 4 in [11])

$$\sup_{0\leq\alpha<1} \|B_{2d+1,\alpha}\|_{L^\infty} = \tilde{B}_{2d}(1/2) = \frac{1}{\pi}\int_{-\infty}^{\infty} \left(\frac{\sin t}{t}\right)^{2d} dt \sim \sqrt{\frac{3}{d\pi}}.$$

In the even case, we get similarly

$$B_{2d+2,1/2}(1/2) = \tilde{B}_{2d+1}(1/2) = \|\tilde{B}_{2d+1}\|_{L^\infty} \geq \tilde{B}_{2d+1}(x^*_{2d+2,\alpha}) = \|B_{2d+2,\alpha}\|_{L^\infty},$$

which gives

$$\begin{aligned}\sup_{0\leq\alpha<1} \|B_{2d+2,\alpha}\|_{L^\infty} &= \tilde{B}_{2d+1}(1/2) \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \left(\frac{\sin t}{t}\right)^{2d+1} dt \sim \sqrt{\frac{6}{(2d+1)\pi}} \sim \sqrt{\frac{3}{d\pi}}.\end{aligned}$$

As a consequence, we get similar formulas for even and odd cases

$$\sup_{0\leq\alpha<1} \|B_{2d+1,\alpha}\|^2_{L^2} \leq \frac{1}{(2d+1)\pi}\int_{-\infty}^{\infty} \left(\frac{\sin t}{t}\right)^{2d} dt \sim \sqrt{\frac{3}{4\pi}} d^{-3/2}, \tag{79}$$

and

$$\sup_{0\leq\alpha<1} \|B_{2d+2,\alpha}\|^2_{L^2} \leq \frac{1}{(2d+2)\pi}\int_{-\infty}^{\infty} \left(\frac{\sin t}{t}\right)^{2d+1} dt \sim \sqrt{\frac{3}{4\pi}} d^{-3/2}. \tag{80}$$

On Figure 3, we plot the numerical value of these quantities to see how well the inequalities (79)–(80) behave for value of $d = 2, \ldots, 9$. The numerical slope of the graphic of $d \to \sup_{0\leq\alpha<1} \|B_{2d+1,\alpha}\|^2_{L^2}$ in logarithmic scale increases (in absolute value) from -1.08 to -1.35. We think that the the asymptotic value of the slope is ≈ -1.5. It is reasonable to infer that the asymptotic slope is indeed -1.5 which, if it is true, will prove that (79-80) are optimal.

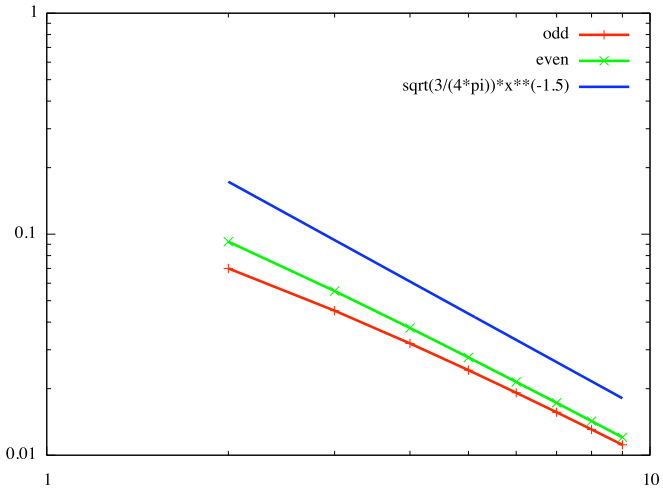

Figure 3: $\sup_{0\leq\alpha<1}\|B_{2d+1,\alpha}\|_{L^2}^2$, $\sup_{0\leq\alpha<1}\|B_{2d+2,\alpha}\|_{L^2}^2$ and $\sqrt{\frac{3}{4\pi}}d^{-3/2}$ versus d (From bottom to top). We observe that the numerical slopes get closer and closer to the theoretical guess $-\frac{3}{2}$.

A.2 Link with Q_i^{p+1}

Q_i^{p+1} has it support in $[x_{i+r+1+k-p}, x_{i+r+1+k}]$, contains $p+2$ points and

$$\int_{x_{i+r+1+k-p}}^{x_{i+r+1+k}} Q_i^{p+1}(x)dx = \frac{1}{p+1}.$$

We then consider $S_{p+1}(t) = Q_i^{p+1}(x_{i+r+1+k-p} + tp\Delta x)$, which has it support in $[0,1]$. We have

$$\int_0^1 S_{p+1}(t)dt = \frac{1}{p(p+1)\Delta x}.$$

On the other hand, $B_{p+1,\alpha}$ contains $p+2$ points, has it support in $[0,1]$ and $\int_0^1 B_{p+1,\alpha}(t)dt = \frac{1}{p+2}$. We deduce that

$$p(p+1)\Delta x S_{p+1}(t) = (p+2)B_{p+1,\alpha}(t),$$

and thus

$$Q_i^{p+1}(x) = \frac{p+2}{p(p+1)\Delta x} B_{p+1,\alpha}\left(\frac{x - x_{i+r+1+k-p}}{p\Delta x}\right).$$

We finally get the optimal bound

$$0 \leq Q_i^{p+1}(x) \leq \frac{p+2}{p(p+1)\Delta x}\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^{p+1} dt \sim \frac{1}{\Delta x}\sqrt{\frac{6}{\pi}}p^{-3/2}. \qquad (81)$$

Remark 25. *The comparison of inequality (81) with the non optimal formula (37) shows that (81) is better by a factor $p^{-\frac{1}{2}}$. Using it in (59), we can improve the estimation of $\|Q_i^{p+1}\|_{L^2}$ by a factor $p^{-1/4}$ and thus obtain (54) by a different method.*

B A Lemma

Lemma 26. *There exists a constant $M > 0$ independent of $p \geq 2$ such that*

$$\left(\int_0^\infty \left(\int_0^\theta |\sin^p(\varphi)| d\varphi \right)^2 \frac{d\theta}{\theta^{2p+2}} \right)^{\frac{1}{2}} \leq \frac{M}{p^{1+\frac{1}{4}}}. \tag{82}$$

Proof. Since $|\sin(\varphi)| \leq \varphi$ for $\varphi \geq 0$ one gets $\int_0^\theta |\sin^p(\varphi)| d\varphi \leq \frac{\theta^{p+1}}{p+1}$. So

$$A := \int_0^\infty \left(\int_0^\theta |\sin^p(\varphi)| d\varphi \right)^2 \frac{d\theta}{\theta^{2p+2}} \leq \frac{1}{p+1} \int_0^\infty \int_0^\theta |\sin(\varphi)|^p d\varphi \frac{d\theta}{\theta^{p+1}}.$$

Therefore

$$A \leq \frac{1}{p+1} \int_0^\infty \left(\int_\varphi^\infty \frac{1}{\theta^{p+1}} d\theta \right) |\sin(\varphi)|^p d\varphi = \frac{1}{p(p+1)} \int_0^\infty \frac{|\sin(\varphi)|^p}{\varphi^p} d\varphi.$$

Without restriction, we assume that $p \geq 2$ so that all integrals are convergent. The last integral is bounded as

$$\int_0^\infty \frac{|\sin(\varphi)|^p}{\varphi^p} d\varphi = \int_0^1 \cdots + \int_1^\infty \cdots \leq \int_0^1 \frac{|\sin(\varphi)|^p}{\varphi^p} d\varphi + \frac{1}{p+1}.$$

Moreover there exists a constant $k > 0$ such that $\sin(\varphi) \leq \varphi - k\varphi^3$ for $0 \leq \varphi \leq 1$. Then

$$\int_0^1 \frac{|\sin(\varphi)|^p}{\varphi^p} d\varphi \leq \int_0^1 (1 - k\varphi^2)^p d\varphi \leq \int_0^1 e^{-pk\varphi^2} d\varphi \leq \int_0^\infty e^{-pk\varphi^2} d\varphi = \frac{1}{2}\sqrt{\frac{\pi}{pk}}.$$

Fnally $A \leq \frac{1}{p(p+1)} \left(\frac{1}{2}\sqrt{\frac{\pi}{pk}} + \frac{1}{p+1} \right)$ from which we deduce the claim. □

References

[1] M. Campos Pinto, M. Mehrenberger, Convergence of an adaptive semi-Lagrangian scheme for the Vlasov-Poisson system, Numerische Mathematik 108 (2008), no. 3, pp. 407-444.

[2] S. Delpino and H. Jourdren, Arbitrary high-order schemes for linear advection and waves equations: application to hydrodynamics and aeroacoustic, Comptes Rendus Acad. Sciences, I, 2006

[3] S. Del Pino, B. Després, P. Havé, H. Jourdren and P. F. Piserchia, 3D finite volume simulation of acoustic waves in the earth atmosphere. Comput. and Fluids 38 (2009), no. 4, 765-777

[4] B. Després, Finite Volume Transport Schemes, Numer. Math. 108 (2008), no. 4, 529-556.

[5] B. Després, Uniform asymptotic stability of Strang's explicit compact schemes for linear advection. SIAM J. Numer. Anal. 47 (2009), no. 5, 3956-3976.

[6] M. Falcone, R. Ferretti, and T. Manfroni, Optimal discretization steps in semi- Lagrangian approximation of first order PDEs, in Numerical Methods for Viscosity Solutions and Applications (Heraklion, 1999), Ser. Adv. Math. Appl. Sci. 59, M. Falcone and C. Makridakis, eds., World Scientific, River Edge, NJ, 2001, pp. 95–117.

[7] W. Hundsdorfer and J. Verwer. Numerical solution of time-dependent advection-diffusion-reaction equations, volume 33 of Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 2003.

[8] A. Iserles and G. Strang, The optimal accuracy of difference schemes, Trans. of the AMS, Vol. 277, 2, 198, 779–803, 1983.

[9] A. Iserles and S.P. Norsett, *Order stars*, Chapman & Hall, 1991.

[10] N. Besse, M. Mehrenberger, Convergence of classes of high-order semi-lagrangian schemes for the Vlasov-Poisson system, Mathematics of computation, 77, 93–123 (2007).

[11] G. Meinardus, H. Morsche, G. Walz, On the Chebyshev Norm of Polynomial B-Splines, Journal of Approximation Theory, 82, 99–122 (1995).

[12] J. Schaeffer, Global existence of smooth solutions to the Vlasov poisson system in three dimensions Communications in Partial Differential Equations Volume 16, Issue 8 & 9, 1991, Pages 1313 –1335

[13] L. Schumaker, *Spline functions: basic theory*, Cambridge University Press, 2007.

[14] G. Strang, Trigonometric polynomials and difference methods of maximum accuracy, J. Math. Phys, 41, 147–520, 1962.

[15] V. Thomée, Stability of difference schemes in the maximum-norm, J. Differential Equations, 1, 273–292, 1965.

[16] V. Thomée, P. Brenner and L. Wahlbin, *Besov spaces and applications to difference methods for initial value problems*, Springer Lecture Notes in Mathematics, 434, Springer-Verlag 1975.